GPO Modifies PURL Referrals Reporting

GPO provides statistical data on a monthly basis that reports the volume of click-throughs by both domain and IP address to online Federal resources through GPO's PURL server. In 2009, Federal depository libraries expressed concerns that the referral statistics were being under-reported. Based on community concerns, GPO evaluated the PURL referral process to determine if abnormalities existed.

After examining the log files, GPO has determined that its PURL referral statistical reporting process was indeed correct based on the definition of a referral. A referral is the URL of a previous item which led to this request; however, after examining the log files and talking with members of the depository library community, a new solution has evolved to report PURL referrals.

As we investigated the PURL system log files, it was determined that a majority of our traffic had no referrer. The referrer field is an optional part of the HTTP request that is sent by a browser to a Web server. There are many reasons why our traffic logs would not contain referral information, including:

 * Requests originate from a proxy server that filters referrals
 * Browsers are not passing referrers or set not to pass referrers by users
 * PURLs are being accessed from browser bookmarks
 * PURLs are being accessed from links within email

Based on our examination of the logs and understanding of the evolution of Internet technologies, GPO has modified its PURL referral parsing script to examine statistics based on the remotehost in order to determine if the request contains either a referrer and/or user-agent. If neither is detected, these requests are considered to be a link validator and ignored. GPO recognizes that there will be ro(bots) that have a referrer and remotehost and will, therefore, be included in our statistical output. We are asking libraries to identify their bots when totaling figures for their institution.

GPO will use the modified PURL referral method for future reporting. Files will be disseminated in comma separated value (csv) format in the **file repository**. Libraries will be responsible for extracting their institution's click-throughs from the domain/IP list contained in the csv.

More information on PURLs can be found in **Linking to Federal Resources Using Persistent Uniform Resource Locators (PURLs)**.